

The relative gains of black students appear on a variety of tests administered to students of different ages in different localities. They appear at ages 9, 13, and 17 in the National Assessment of Educational Progress (Figure IV-5); in the SAT; in a nationally representative comparison of high school seniors in 1971 and 1979; in grades 3, 6, and 9 in the North Carolina state assessment program; among ninth graders in the Texas state assessment program; and in test data from some local education agencies, such as Cleveland, Houston, and Montgomery County (Maryland). ^{25/}

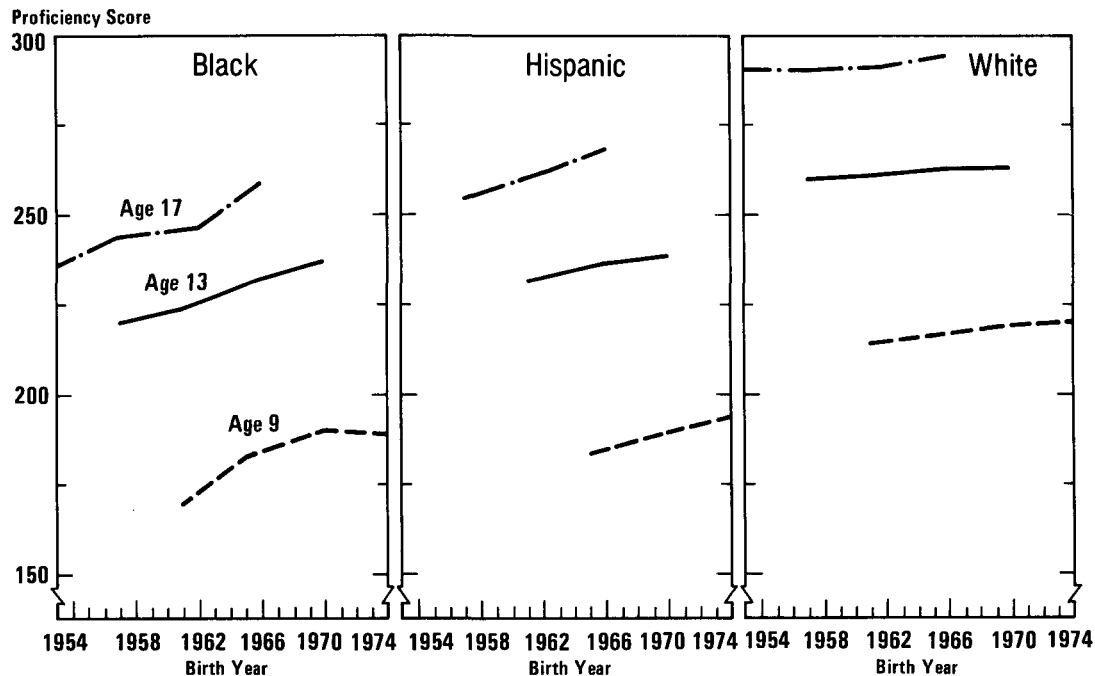
The SAT data suggest that part of the convergence of black and non-minority scores resulted from the decline ending earlier among black than among nonminority students. The convergence of scores continued during the period of the general upturn, however, as black students gained more rapidly than did nonminority students.

Although this shrinking of the gap has been small relative to the average differences between black and nonminority students, the rate of change has been appreciable. For example, over the past nine years, the gap between black and nonminority students on the SAT has shrunk at an annual rate roughly comparable to the average rate of the total SAT decline--a change that few people would label insignificant. On the National Assessment, the average black student's mathematics score was a third below the nonminority average in 1972 but a fourth below that in 1981.

-
25. National Assessment of Educational Progress (NAEP), *Three National Assessments of Science*; NAEP, *Three National Assessments of Reading*; NAEP, *The Reading Report Card*; NAEP, *The Third National Mathematics Assessment*; and NAEP, *Mathematical Technical Report: Summary Volume*; *College Board Data Show Class of '85 Doing Better on SAT, Other Measures of Educational Achievement* (New York: The College Board, September 23, 1985); Rock and others, *Factors Associated with Decline of Test Scores*, Tables D-1, D-2, and D-3; Nancy W. Burton and Lyle V. Jones, "Recent Trends in Achievement Levels of Black and White Youth," *Educational Researcher*, vol. 11 (April 1982), pp. 10-14, 17; Montgomery County (Maryland) Public School District, "MCPS Test Results by Racial/Ethnic Groups, 1977-1982," unpublished paper; Marian Kilbane-Flash, personal communication, March 1984; and Houston Independent School District, unpublished tabulations.

On the other hand, scores on the ACT are not entirely consistent with this pattern. The gap between black and other students on the ACT composite has narrowed since 1970, but only slightly, and the trend has been highly erratic from year to year. In addition, the trend varies among subjects; the gap narrowed in social studies, for example, but grew slightly in mathematics. This partial inconsistency with the patterns evident in other tests is discussed further in Appendix E.

Figure IV-5.
Trends in Average Reading Proficiency for White, Black, and
Hispanic Students, by Birth Year



SOURCE: National Assessment of Educational Progress, *The Reading Report Card* (Princeton: NAEP/Educational Testing Service, 1985), Data Appendix.

It is likely, but not certain, that this narrowing of the gap will continue to appear in some test data for several years. The NAEP data show the most rapid convergence accompanying the birth cohorts of the mid-1960s as they pass through school--appearing at age 9 in the early 1970s and at age 17 in the early 1980s. Some narrowing, however, appeared at least as late as the birth cohorts of the late 1960s and perhaps as late as those of the early 1970s.^{26/} This pattern would suggest further convergence between black and nonminority scores on high school tests for several years. On the other hand, the SAT is inconsistent with this pattern; the relative gains of black students on that test ended in 1981 and 1982.

26. National Assessment of Educational Progress, *The Reading Report Card*, Figure 3.2.

Despite these changes, the gap between the average scores of black and nonminority students remains striking. On the SAT, for example, the average black student's score in 1975 corresponded roughly to the 11th and 12th percentiles among nonminority students on the mathematics and verbal scales, respectively. In 1984, the average black scores had risen to about the 16th percentile among nonminority scores on both scales.^{27/} While some other tests show smaller average differences than the SAT, the gap nonetheless remains large by virtually any measure.

Hispanic Students. In national samples, Hispanic students on average show substantially lower levels of achievement than nonminority students, though somewhat higher achievement than blacks. In recent years, the average achievement of Hispanic students, like that of blacks, has risen relative to that of nonminority students.

Generalizations about achievement trends among Hispanic students, however, are subject to important qualifications. First, the relevant data are more limited than in the case of blacks. More important, the term "Hispanic" subsumes many groups differing in culture of origin, length of residence in the United States, relative fluency in and use of English and Spanish, and other factors that presumably affect educational performance. Thus, trends among Hispanic students as a whole provide only suggestions of trends that might be occurring in more specific groups that are often the targets of specific educational programs--such as children with limited proficiency in English, or the children of migrant farm workers.

With those qualifications in mind, the relative improvement of Hispanics' achievement is apparent in the NAEP reading and mathematics assessments (see Figure IV-5), in the SAT, in the Texas state-wide assessment of ninth grade students, and in a comparison of nationally representative samples of high school seniors in 1971 and 1979 (the NLS and HSB comparison).^{28/} This trend appears not to be limited to one Hispanic group. Relative gains appear among both Mexican-American and Puerto Rican students on the SAT and among both Mexican Americans and "other Hispanics" in the NLS and HSB comparison, although the improvement among Mexican Americans is in several instances greater.^{29/} The annual

27. These estimates are based on nonminority within-group standard deviations in 1983-1984 reported in Solomon Arbeiter, *Profiles of College-Bound Seniors, 1984* (New York: The College Entrance Examination Board, 1984), p. 81.

28. Rock and others, *Factors Associated with Decline of Test Scores*, Appendix D. In this instance, however, the differences in the trends shown by Hispanics and nonminority students are slight.

29. The changes in these Hispanic subgroups in the NLS and HSB comparison, however, appear somewhat unstable and are statistically not significantly different from no change.

SAT data suggest that among Hispanics--as among blacks--the achievement decline ended a few years earlier than it did among nonminority students.

DIFFERENCES IN TRENDS AMONG TYPES OF SCHOOLS AND COMMUNITIES

While the achievement decline was pervasive, it has not been entirely uniform among different types of communities and schools. This section discusses the relative trends in three specific types of schools and communities about which data are available:

- o Disadvantaged urban communities;
- o Schools with different concentrations of ethnic minorities; and
- o Private schools.

Disadvantaged Urban Communities

Since 1970, 9- and 13-year-olds in disadvantaged urban communities gained ground relative to the nation as a whole on the NAEP mathematics and reading assessments (see Tables IV-1 and IV-2).^{30/} In contrast, 17-year-olds in disadvantaged urban communities showed no relative gains in mathematics, and their small relative gains in reading occurred entirely between 1979 and 1983. In two instances--in reading at age 9, and in mathematics at age 13--more than a third of the gap between disadvantaged-urban communities and the nation as a whole was overcome since the early 1970s.^{31/}

-
30. For a school to be defined as "disadvantaged urban," it had to be located within either the city limits or the urban fringe of a city of at least 200,000 people (or twin or triplet cities with combined populations over 200,000); and it had to serve a community that had unusually few managerial and professional personnel and atypically many unemployed adults and adults on welfare. The latter criterion was implemented through four steps: asking the principal of each school to estimate the proportion of students whose parents fell into those categories; summing the percentages on welfare and unemployed; subtracting the percentage professional or managerial; and selecting the schools that constituted the top 10 percent on the resulting index. (Westat Corporation, unpublished NAEP documentation).
31. In the case of mathematics, however, the amount by which the gap closed can be considered only approximate, for the 1972 average scores are only estimates. See footnote A, Table IV-1.

TABLE IV-1. AVERAGE MATHEMATICS ACHIEVEMENT IN DISADVANTAGED URBAN COMMUNITIES AND IN THE NATION, NAEP, 1972-1981
(Average percent of items correctly answered)

Group	1972 (Estimated) a/	1977	1981	Percent Change 1972-1981
Age 9				
Nation	56.7	55.4	56.4	-1
Disadvantaged Urban	41.9	44.4	45.5	9
Nation Minus Disadvantaged Urban	14.8	11.0	10.9	-26
Age 13				
Nation	58.6	56.6	60.5	3
Disadvantaged Urban	41.5	43.5	49.3	19
Nation Minus Disadvantaged Urban	17.1	13.1	11.2	-35
Age 17				
Nation	64.0	60.4	60.2	-6
Disadvantaged Urban	51.5	45.8	47.7	-7
Nation Minus Disadvantaged Urban	12.5	14.6	12.5	0

SOURCES: CBO calculations based on National Assessment of Educational Progress, *The Third National Mathematics Assessment: Results, Trends, and Issues* (1983), Tables 5.1 and 5.2; and *Mathematics Technical Report: Summary Volume* (1980), Tables 2, 3, and 4.

- a. These estimates for 1972 differ from published NAEP results for the 1972 assessment. The published results for that year are based either on the 1972 item pool or on the items used in both 1972 and 1977, while the trend results comparing the 1977 and 1981 assessments reflect items used in both the 1977 and 1981 assessments. In order to circumvent the large disparities in the item sets, 1972 results were estimated here by adjusting the 1977 results (on the items used in 1977 and 1981) by the 1972-to-1977 change (on the items used in 1972 and 1977).

TABLE IV -2. AVERAGE READING ACHIEVEMENT IN
DISADVANTAGED URBAN COMMUNITIES AND
IN THE NATION, NAEP, 1970-1983
(Average proficiency scores)

Group	1970	1974	1979	1983	Percent Change 1970-1983
Age 9					
Nation	207	210	214	213	3
Disadvantaged Urban	178	185	186	194	9
Nation Minus Disadvantaged Urban	29	25	28	19	-34
Age 13					
Nation	254	255	257	258	2
Disadvantaged Urban	232	229	242	240	3
Nation Minus Disadvantaged Urban	22	26	16	18	-18
Age 17					
Nation	284	285	285	288	1
Disadvantaged Urban	259	261	258	266	3
Nation Minus Disadvantaged Urban	25	24	26	22	-12

SOURCES: National Assessment of Educational Progress, *The Reading Report Card*, Data Appendix.

NOTE: Details might not add to totals because of rounding.

Schools With High or Low Concentrations of Minority Students

Although information on the relative trends in high- and low-minority schools is limited, such data as are available suggest that, relative to the nation as a whole, high-minority schools have gained in achievement while low-minority schools have lost ground. While the available analyses of these data do not clarify whether the gains of minority students have been larger or smaller in high-minority schools, they do indicate that the relative gains of minority students as a group cannot be attributed entirely to improved performance of those attending low-minority schools. At all ages, mathematics gains between the last two National Assessments (1977 and 1981) were several times as large in schools that had minority enrollments of at least 40 percent than in other schools (see Table IV-3). Similarly, in a comparison of the HSB and NLS test results, seniors in low-minority schools -- defined as at least 90 percent nonminority -- showed, on average, larger declines from 1972 to 1980 than did seniors in other schools. In the case of vocabulary, the decline in low-minority schools was 83 percent larger than in other schools. The difference was about half that size in mathematics, and a fourth in reading. ^{32/}

Private Schools

The achievement decline occurred among high school students in private as well as public schools. Moreover, it appears to have been nearly as large among private school students in reading and vocabulary, although somewhat smaller in mathematics (if tests of reading, vocabulary, and mathematics administered to seniors during the last half of the decline are an adequate indication). ^{33/} Beyond that, very little can be said about the relative trends among private school students, because of the extremely sparse data. For example, whether the upturn in achievement found in public school and nationally representative data--the latter of which is dominated by the far more numerous public school students--occurred in private schools as well is not yet known.

32. Rock and others, *Factors Associated with Decline of Test Scores*, Appendix D.

33. Ibid.

TABLE IV-3. AVERAGE MATHEMATICS ACHIEVEMENT IN
HIGH-MINORITY AND LOW-MINORITY SCHOOLS,
NAEP, 1977 AND 1981 (Average percent
of items correctly answered)

Group	1977	1981	Percent Change 1981-1977
Age 9			
Nation	55.4	56.4	1.8
40 Percent or More Minority	46.4	48.8	5.2
Less than 40 Percent Minority	57.6	58.6	1.7
Age 13			
Nation	56.6	60.5	6.9
40 Percent or More Minority	45.5	53.6	17.8
Less than 40 Percent Minority	59.6	62.4	4.7
Age 17			
Nation	60.4	60.2	-0.3
40 Percent or More Minority	47.5	52.3	10.1
Less than 40 Percent Minority	62.4	62.4	0.0

SOURCE: National Assessment of Educational Progress, *The Third National Mathematics Assessment: Results, Trends, and Issues* (1983), Table 5.2.

The SAT decline was found among both private- and public school students.^{34/} Since the selection changes that contributed to the SAT decline might have been very different among private school students, however, a comparison of the size of the SAT decline in the two groups of students would be risky.

34. Advisory Panel on the Scholastic Aptitude Test Score Decline, *On Further Examination*, p. 20.

APPENDIXES





APPENDIX A

DESCRIPTION OF MAJOR DATA SOURCES

This Appendix briefly describes the most important data sources used in the text and in other appendixes. These sources are:

- o Two college-admissions tests--the Scholastic Aptitude Test (SAT) and the American College Testing Program (ACT) tests;
- o The National Assessment of Educational Progress (NAEP);
- o The test data from two nationally representative studies of high school students--the National Longitudinal Study of the High School Seniors Class of 1972 (NLS) and the High School and Beyond study (HSB); and
- o Annual statewide test data from Iowa.

THE SCHOLASTIC APTITUDE TEST

The Scholastic Aptitude Test (SAT), sponsored by the College Board and administered by the Educational Testing Service, is intended to aid colleges in selecting students for admission. It is perhaps the single best known test in the United States and has figured prominently in discussions of achievement trends for a decade or more.

The SAT is taken by a large number of students, but they constitute a clearly nonrepresentative group. Students taking the test are predominantly those intending to attend college, have higher levels of achievement than does the student body as a whole, and are concentrated in certain geographic regions. In the 1984-1985 school year, the SAT was taken by nearly one million high school students, representing over a third of all graduates and about two-thirds of college-bound graduates.^{1/} Nonetheless, it was the

-
1. The College Entrance Examination Board, *National College-Bound Seniors, 1985* (New York: The College Board, 1985). The number of high school graduates in the 1984-1985 school year, excluding high school equivalency credentials, has been projected to be about 2.6 million. National Center for Education Statistics, *Projections of Education Statistics to 1990-91* (Washington, D.C.: NCES, 1982), Table 15.



principal college admissions test in only 22 states, which were primarily in the east and on the west coast. ^{2/}

The SAT consists of two tests, one mathematical and one verbal.^{3/} The verbal test consists of analogies, antonyms, sentence completions, and reading passages.^{4/} The mathematics test consists of a variety of problems in arithmetic reasoning, algebra, and geometry that are intended to "require as background mathematics typically taught in grades one through nine" but to "depend less on formal knowledge than on reasoning."^{5/}

The SAT is designed to predict achievement in college, not to directly assess achievement in secondary schools. Accordingly, the test has been validated primarily by documenting that students scoring higher on the test tend to have higher grades in college.^{6/} In contrast, tests intended to assess students' current levels of mastery are typically validated by showing that students scoring higher on the test in question tend to score higher on some other measure of current achievement, such as teachers' evaluations or other achievement tests.^{7/}

Although the SAT is designed to be a predictor of college performance and was neither intended nor validated as an achievement test, it has often been used as an index of achievement--despite strong objections from the

-
2. *State Education Statistics: State Performance Outcomes, Resource Inputs, and Population Characteristics, 1982 and 1984* (Washington, D.C.: U.S. Department of Education, January 1985).
 3. A third scale, the "Test of Standard Written English," was first added on an experimental basis in the mid-1970; it is not discussed in this paper.
 4. The College Board, *College-Bound Seniors*.
 5. Advisory Panel on the Scholastic Aptitude Test Score Decline, *On Further Examination* (New York: The College Board, 1977), p. 9.
 6. Hunter M. Breland, *Population Validity and College Entrance Measures* (New York: The College Board, 1979). It is well established that high SAT scores are associated with higher grades early in college. The extent to which the SAT provides information about likely college performance above and beyond that provided by other indices such as high school grades is a matter of some disagreement. That issue, however, is not germane to the use of SAT scores in this paper. (See, for example, James Crouse, "Does the SAT Help Colleges Make Better Selection Decisions?" *Harvard Educational Review*, vol. 55, May 1985, pp. 195-219; and George H. Hanford, "Yes, the SAT Does Help Colleges," *Harvard Educational Review*, vol. 55, August 1985, pp. 324-331.)
 7. See, for example, Science Research Associates, *SRA Achievement Series, Technical Report #3* (Chicago: SRA, 1981).

College Board.^{8/} For example, much of the public debate about declining achievement focused at least in part on the SAT, and the annual compilation of state education statistics by the U.S. Department of Education calls the test a "performance outcome" (rather than a "predictor of performance").^{9/}

The SAT is administered several times each year, and the scores obtained in each year are equated, so that any given score should reflect approximately the same level of skill in any year. Annual publications provide detailed tabulations of the scores of the test-taking group as a whole and of a variety of subgroups, such as males, females, and ethnic groups. Data on student characteristics such as these are mostly based on a Student Descriptive Questionnaire (SDQ) completed by students, and the information is therefore subject to distortions stemming from both non-response and various kinds of reporting errors.

Data on the SAT extend back longer than those on most other tests, but the long-term data used in this paper are subject to several inconsistencies. Current tabulations by the College Board reflect only the most recent test taken by students who also completed the SDQ--about 90 percent of all SAT candidates.^{10/} Average scores from the 1966-1967 through 1970-1971 school years are College Board estimates of the averages that would have been obtained if such tabulations had been made for those years. Data from the 1956 through 1965 school years are based on the average of all scores, which includes multiple scores by those taking the SAT more than once.^{11/} The published data on these averages of all scores were adjusted by subtracting from them the slight difference in 1966 between that average and the average based on only the most recent of each individual's scores. Trend data on the proportion of SAT scores above specific thresholds were subject to a similar discontinuity and were similarly adjusted, but in that case the adjustment was based on the average discrepancy in averages over

-
8. See, for example, statement by Daniel B. Taylor, Senior Vice President, the College Board, before the House Subcommittee on Elementary, Secondary, and Vocational Education, Committee on Education and Labor, January 31, 1984.
 9. *State Education Statistics: State Performance Outcomes, Resource Inputs, and Population Characteristics, 1982 and 1984* (Washington, D.C.: U.S. Department of Education, January 1985).
 10. The College Board, *College-Bound Seniors, 1985*, p. 4.
 11. Hunter M. Breland, *The SAT Score Decline: A Summary of Related Research* (New York: The College Board, 1976), Table 1.

a four-year period (1971 through 1974) for which both averages were available. 12/

THE AMERICAN COLLEGE TESTING PROGRAM TESTS

The American College Testing Program (ACT) tests, like the SAT, are intended as an aid in selecting students for admission to college. The ACT tests were taken by about 739,000 high-school students in the class of 1984-1985--over a fourth of all graduates. Although the ACT battery is taken by fewer students than is the SAT, it is the predominant college-admissions test in 28 states--primarily in the Midwest, the western mountain states, and parts of the Southeast. 13/

Although also intended to predict success in post-secondary education, the ACT is conceptually distinct from the SAT and is in some senses intended to be more of a test of achievement. The ACT is more "curriculum based" than is the SAT, relying on both reasoning ability and knowledge of subject-matter fields. Despite its intentional reliance on subject-matter knowledge, however, the ACT contains many "analytical, problem-solving exercises and few measures of narrow skills." 14/

The ACT battery consists of subject-matter tests in English, mathematics, social studies, and natural science, yielding four subject-specific scores as well as a composite score. The English test is a test of usage, tapping skills such as grammar, sentence structure, and paragraph organization. The mathematics test is dominated by questions on arithmetic and algebraic reasoning, geometry, and intermediate algebra, but a fourth of the test is devoted to arithmetic and algebraic operations, number concepts, and advanced topics. The social studies test includes aspects of history, government, anthropology, sociology, psychology, and economics. The

12. Ibid., Table 5.

13. American College Testing Program, *National ACT Assessment Results, 1984-1985: Executive Summary* (Iowa City: ACT, 1985); U.S. Department of Education, *State Education Statistics*.

14. *Content of the Tests in the ACT Assessment* (Iowa City: American College Testing Program, undated).

natural sciences test is about evenly divided between chemistry, physics, other physical sciences, and biology. 15/

The ACT is reported and equated annually. Trend data reflecting subgroups of students are available but are less extensive than those available for the SAT.

The long-term ACT trend data used in this paper are subject to one inconsistency. Scores from 1969 on are taken from internally consistent tabulations published by ACT. 16/ Earlier data are adapted from tabulations that differ from the more recent data in including scores from "residual" testing of students on college campuses, who have lower average scores than those taking the test before college. 17/ These earlier averages were adjusted by adding to them the small difference in 1969 between them and the averages consistent with later data. 18/

THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

The National Assessment of Educational Progress (NAEP) is a critical indicator of achievement trends, for it alone among current data sources provides repeated testing of nearly representative samples of the national student population.

Before the NAEP was begun, available data often provided an indication of achievement patterns and trends in smaller areas--that is, in schools, districts, or occasionally states. But variations in assessment methods from one jurisdiction to another precluded using these data as an unambiguous indicator of achievement across the entire nation.

In contrast, the NAEP was designed to be a measure of the performance of the nation's elementary and secondary educational system as a whole. It was not intended to duplicate the assessment mechanisms already in place. For example, it was intended to assess relatively general levels of

-
15. American College Testing Program, *Content of the Tests*.
 16. For example, *American College Testing Program, National Trend Data for Students Who Take the ACT Assessment* (Iowa City: ACT, undated).
 17. James Maxey, American College Testing Program, personal communication, April 1984.
 18. The unadjusted earlier data are in L. A. Munday, *Declining Admissions Test Scores* (Iowa City: American College Testing Program, 1976), Table 3.

knowledge, and it was not designed to differentiate among individuals. It was to supplement those other measures by providing a consistent, broad measure of the achievement of a largely representative sample of the nation's youth that would be periodically repeated. 19/

Since 1969, the NAEP has provided periodic testing of 9-, 13-, and 17-year-old students in 10 subject areas. The intervals between assessments in any subject area typically have ranged from three to five years. The best known assessments are in the areas of reading, writing, mathematics, science, and social studies. 20/

Although the NAEP is nearly representative of students nationwide, it excludes several important groups. In most instances, the NAEP has tested only those individuals still in school. 21/ In the case of 17-year-olds, this practice leads to results that are probably quite different from those that would be obtained if all 17-year-olds were tested, since dropouts are numerous in that age group and tend to be low achievers. The overall average score is thus higher than it would be, and comparisons between groups (ethnic groups, regions, and so on) reflect differences in dropout rates as well as achievement differences in the entire age cohort. In addition, handicapped students and those with limited proficiency in English are excluded from testing, although the definition of those categories can vary somewhat from one participating school to another. Both of these exclusions are germane to the assessment of trends, since the period over which the NAEP has been conducted saw the passage of the Education of the Handicapped Act (which most likely increased the number of handicapped students in regular school programs markedly) and rapid immigration from Latin America and Asia. Finally, participating schools have some discretion to exclude other students who cannot be assessed properly. 22/

-
19. *Director's Report to the Congress on the National Assessment of Educational Progress* (Washington, D.C.: National Institute of Education, December 1982).
 20. At various times, the National Assessment has included tests of other groups and subjects that are not considered here.
 21. Brief descriptions of the NAEP sampling procedure are provided in a number of publications. See, for example, *National Assessment of Educational Progress, Mathematical Technical Report: Summary Volume* (Denver: NAEP/Education Commission of the States, 1980), Chapter 1.
 22. Lawrence Rudner, Office of Educational Research and Improvement, U.S. Department of Education, personal communication, December 1985.

The NAEP tests are designed to assess a range of skills varying in difficulty. In mathematics, for example, the easiest items tap recall of factual information and simple arithmetic computation. More difficult items require an ability to manipulate algebraic expressions, to comprehend and explain mathematical relationships, and to apply skills in solving problems. ^{23/}

For the purposes of this paper, the principal advantages of the NAEP are its nearly representative sampling, its diversity of subject areas and levels of skills, and a considerable amount of background information. A variety of characteristics of students, schools, and communities were ascertained through student, teacher, and school questionnaires. These data permit comparisons of trends, for example, among ethnic groups, geographic regions, and schools with high and low minority enrollments.

These advantages are mitigated, not only by the time intervals between assessments, but also by the forms in which data were presented and the lack of formal equating of scores from one assessment to another. Until recently, scores were generally only reported as the percentage of items answered correctly--a scaling that has some intuitive appeal but one that poses serious problems in gauging trends and, especially, in comparing trends among groups. ^{24/} In addition, information on the standard deviation of average scores was often not reported or retained, limiting the extent to which the severity of trends could be quantified and compared with that on other tests. Beginning with the most recent assessment of reading, these problems have in large part been solved, but most of the trend data remain in the original form. Scores were also not formally equated until recently, posing problems in the interpretation of trends that were compounded by periodic alteration of the content of the tests. A frequent, but not fully adequate, response to this problem in the published NAEP data was to base comparisons only on items shared by adjacent assessments.

THE NATIONAL LONGITUDINAL SURVEY AND HIGH SCHOOL AND BEYOND

Two nationally representative longitudinal studies of high school students--the National Longitudinal Study of the High School Seniors Class of 1972

23. See, for example, National Assessment of Educational Progress, *Changes in Mathematical Achievement, 1973-78* (Denver: NAEP/Education Commission of the States, 1979).

24. See Chapter II.

(NLS) and the High School and Beyond study (HSB)--provide comparative information on the achievement of seniors in the 1971 and 1979 school years. 25/

Both studies included a variety of cognitive tests, of which three that were administered in both years--vocabulary, reading, and mathematics--can be considered measures of achievement. 26/ The reading and vocabulary tests were identical in the two studies; in mathematics, about half of the items were identical, a fourth were altered in relatively minor respects, and the remainder were new.

In one recent study, the scores on the NLS and HSB tests in those three subject areas were equated, providing an indication of changes in performance over the eight years. 27/ All comparisons of the NLS and HSB in this paper are drawn from that study.

Information is available in the NLS and HSB about a considerable number of important student, school, and community variables, making possible both comparisons of achievement changes in different groups and estimation of the effects of population changes (such as trends in the ethnic composition of the school-age population) on average test scores. This information is derived from school records, school questionnaires, and teacher questionnaires, as well as from student self-reports, which increases the validity of some of the information compared with that obtained solely through student questionnaires. Moreover, in some instances, it permits information from one source to be confirmed by comparing it with that from another.

The usefulness of the NLS and HSB for analyzing achievement trends is limited by several factors, however. The absence of earlier, comparable

-
25. The NLS and HSB tests were administered in the springs of 1972 and 1980, and most discussions of them refer to those calendar years. In order to be consistent with the treatment of other tests, however, this paper refers instead to the school years in which the tests were administered.
 26. Other tests tapped basic cognitive skills but could not be considered measures of achievement. For example, a mosaic comparisons test was included in 1972 as an index of "perceptual speed and accuracy." For a brief description of the two test batteries, see Donald A. Rock, Ruth B. Ekstrom, Margaret E. Goertz, Thomas L. Hilton, and Judith Pollack, *Factors Associated with Decline of Test Scores of High School Seniors, 1972 to 1980* (Washington, D.C.: Center for Statistics, U.S. Department of Education, 1985) Chapter II.
 27. Donald Rock and others, *Factors Associated with Decline of Test Scores*.